BioUML
integrated informational platform for biomedical research

Fedor Kolpakov

Biosoft.Ru, Ltd.
Institute of Systems Biology, Ltd.

Novosibirsk, Russia
BioUML platform

- BioUML is an open source integrated platform for systems biology that spans the comprehensive range of capabilities including access to databases with experimental data, tools for formalized description, visual modeling and analyses of complex biological systems.
- Due to scripts (R, JavaScript) and workflow support it provides powerful possibilities for analyses of high-throughput data.
- Plug-in based architecture (Eclipse run time from IBM is used) allows to add new functionality using plug-ins.

BioUML platform consists from 3 parts:
- BioUML server – provides access to biological databases;
- BioUML workbench – standalone application.
- BioUML web edition – web interface based on AJAX technology;
Plug-in based architecture

A **plug-in** is the smallest unit of BioUML workbench function that can be developed and delivered separately into BioUML workbench. A plug-in is described in an XML manifest file, called plugin.xml. The parsed contents of plug-in manifest files are made available programmatically through a **plug-in registry** API provided by Eclipse runtime.

- **extension points** are well-defined function points in the system where other plug-ins can contribute functionality.

- **extension** is a specific contribution to an extension point. Plug-ins can define their own extension points, so that other plug-ins can integrate tightly with them.
BioUML workbench
Integrated model of apoptosis. Fas/LCD95 signaling pathway.

**ID:** PRT001953

**Title:** Fas

**Complete name:** CMP0219 PRT001953

**Role:** VariableRole

```
"CMP0219.PRT001953"=50.0 null;
```
BioUML main features

• Supports access to main biological databases:
  – *catalogs*: Ensembl, UniProt, ChEBI, GO…
  – *pathways*: KEGG, Reactome, EHMN, BioModels, SABIO-RK, TRANSPATH, EndoNet, BMOND…

• Supports main standards used in systems biology:
  SBML, SBGN, CellML, BioPAX, OBO, PSI-MI…

• database search:
  – full text search using Lucene engine
  – graph search

• graph layout engine

• visual modeling:
  – simulation engine supports (ODE, DAE, hybrid, 1D PDE);
  – composite models;
  – agent based modeling;
  – parameters fitting;

• genome browser (supports DAS protocol, tracks import/export);

• data analyses and workflows – specialized plug-ins for microarray analysis, integration with R/Bioconductor, JavaScript support, interactive script console.
Main platforms for bioinformatics and BioUML

**Taverna**
- standalone application
- powerful workflows

**BioUML platform**
- web interface
- collaborative research
- genome browser

**R/Bioconductor**
- scripts
- statistics, plots

**Galaxy**
- workflows, web interface
- collaborative research
- genome browser

**BioClipse**
- Eclipse plug-in based architecture
- chemoinformatics
Main platforms for bioinformatics and BioUML

**Taverna**
- standalone application
- powerful workflows

**Galaxy**
- workflows, web interface, collaborative research, genome browser

**R/Bioconductor**
- scripts, statistics, plots
  - + systems biology
    - • visual modelling
    - • simulation
    - • parameters fitting
    - • ...
  - + chat for on-line consultations

**BioUML platform**
- web interface, collaborative research, genome browser
- Eclipse plug-in based architecture, chemoinformatics

**BioClipse**
- Eclipse plug-in based architecture, chemoinformatics

---

**BioUML platform**
- visual modelling
- simulation
- parameters fitting
- ...
- chat for on-line consultations
BioUML
web edition
Routes to explain gene regulation and functions

- Load data
- Normalize data
- Detect differentially expressed genes
- Discover functional enrichment
- Identify master regulators in networks
- Search for putative TFBS

Create your own workflow

User description is not available
Icons for local and remote databases

Local database
- local database, available for reading and writing

Remote database
- public remote database, read only
- public remote database, available for reading and writing

before log-in
- remote public database, requires log-in for writing
- remote protected read only database, requires log-in for reading
- remote protected database, requires log-in for reading and writing

after log-in
-
Text search:
universal full text search engine
based on Apache Lucene technology
ID: Diagrams
Complete name: databases/Biopath/Diagrams

<table>
<thead>
<tr>
<th>#</th>
<th>ID</th>
<th>description</th>
<th>components</th>
<th>statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>databases/Biopath/Diagrams/DGR0025</td>
<td>IKK(i), IKK-alpha(i), IKK-beta(i), IKK-g...</td>
<td>components:17, reactions:0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>databases/Biopath/Diagrams/DGR0027</td>
<td>antioxidant enzymes, proinflammatory cytokines...</td>
<td>components:74, reactions:0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>databases/Biopath/Diagrams/DGR0029</td>
<td>transient checkpoint arrest, cell cycle...</td>
<td>components:7, reactions:0</td>
<td></td>
</tr>
</tbody>
</table>
ID: PRT000063
Title: AP-1
Comment: An alteration of redox balance in favour of oxidation triggers intracellular signalling such as the activation of the stress kinases (cJun, JNK, p38 and etc) which leads to transcription factor activation and subsequent transcription of pro-inflammatory genes.
FasL/CD95L/APO-1 signaling

Input: Fas/CD95/APO-1
Output: caspase-8

FasL
Fas
Fas:FasL
Fas:FasL Trimer
AP-1
k2_1
k2_2
k2_3
FADD

ID: PRT000063
Title: AP-1
Complete name: CMP0219.PRT000063
Metaphor: biological systems reconstruction as solitaire game

Desk – BioUML editor

Solitaire – biological pathway

Cards – biological objects (genes, proteins, lipids, etc.)

Pack of cards – different biological databases
Systems Biology

Computational modeling

Experimental data

Analysis

Experimental hypothesis validation
Visual modeling
## Pane: Simulation

### Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output dir</td>
<td>.out</td>
</tr>
<tr>
<td>Initial time</td>
<td>0.0</td>
</tr>
<tr>
<td>Completion time</td>
<td>100.0</td>
</tr>
<tr>
<td>Time increment</td>
<td>1.0</td>
</tr>
<tr>
<td>Absolute tolerance</td>
<td>1.0E-8</td>
</tr>
<tr>
<td>Relative tolerance</td>
<td>1.0E-5</td>
</tr>
<tr>
<td>solverName</td>
<td>JVode</td>
</tr>
</tbody>
</table>

### Needed Plot Simulation

- DormandPrince
- Euler
- Imex
- JVode
ID: z
Title: z
Complete name: databases/BiomedModels/Diagrams/BIOMD0000000006.xml/cell/z
Role: VariableRole: "$cell.z"=0.0 substance;
de=DataElement[z] class=class bioxml.model.Compartment
Attributes:
sbgn:entityType: unspecified
xmlElementType: entity
Online SBML Test Suite

The SBML Test Suite allows you to evaluate the degree and correctness of SBML support implemented in SBML-compatible software. The system currently supports specifications of SBML up through Level 2 Version 4 Release 1. This page is the interface to the online version of the Suite; it allows you to upload test results and have them evaluated by our server. (A standalone version of the Test Suite was previously released in alpha version form and is currently being updated—stay tuned for its release.)

There are three steps to using the online interface:

1. **Select and download test cases.** You can download all cases or select a subset using the interface provided on our online test case selection page.

2. **Run simulations** of the models in the software package you are testing, and collect the results. How you run the cases is up to you and the software you are testing.

3. **Upload the simulation results.** This online service will compare them to the expected results and provide you with a report of the outcome.

Limitations: (1) This online service works only with the semantic portion of the SBML Test Suite. It does not evaluate the results of syntactic tests available with the full SBML Test Suite. (2) Only HTTP is supported as the protocol for uploading results; other protocols such as https://, ftp://, etc. are not supported.
Summary

<table>
<thead>
<tr>
<th>Tests</th>
<th>Successful</th>
<th>Failed</th>
<th>Errors</th>
<th>Needs tuning</th>
<th>CSV error</th>
<th>Result differs</th>
<th>Problem is stiff</th>
<th>Compilation error</th>
<th>Success rate</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>927</td>
<td>927</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100.0%</td>
<td>855</td>
</tr>
</tbody>
</table>

Legend:
Failed - an exception has occurred.
Errors - simulation results significantly differ from the known ones.
Needs tuning - relative error is not small enough.
CSV error - original CSV data is missing or cannot be parsed.
Result differs - some variable or time point is missing in simulation engine output.
Problem is stiff - Problem is too stiff for this solver.
Compilation error - Could not compile generated model file.

<table>
<thead>
<tr>
<th>Name</th>
<th>Tests (s)</th>
<th>Successful</th>
<th>Failed</th>
<th>Errors</th>
<th>Needs tuning</th>
<th>CSV error</th>
<th>Result differs</th>
<th>Problem is stiff</th>
<th>Compilation error</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EventWithDelay</td>
<td>111</td>
<td>111</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>AlgebraicRule</td>
<td>81</td>
<td>81</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>AssignmentRule</td>
<td>181</td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>InitialAssignment</td>
<td>138</td>
<td>138</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
</tbody>
</table>
Reports (templates)
CD95L module

Identifier: Int_CD95L signaling
Type: Pathway simulation

Integrated model of apoptosis. FasL/CD95 signaling pathway.

Compartments

<table>
<thead>
<tr>
<th>№</th>
<th>ID</th>
<th>Name</th>
<th>Initial value</th>
<th>Units</th>
<th>Comment</th>
<th>Database references</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CMP0219</td>
<td>Cell</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Species

<table>
<thead>
<tr>
<th>№</th>
<th>ID</th>
<th>Name</th>
<th>Compartment</th>
<th>Initial value</th>
<th>Units</th>
<th>Boundary condition</th>
<th>Comment</th>
<th>Database references</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PRT001953</td>
<td>Fas</td>
<td>Cell</td>
<td>50.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>PRT001964</td>
<td>FasL</td>
<td>Cell</td>
<td>0.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>PRT001974</td>
<td>caspase-8</td>
<td>Cell</td>
<td>0.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PRT002331</td>
<td>procaspase-8</td>
<td>Cell</td>
<td>25.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PRT002333</td>
<td>FADD</td>
<td>Cell</td>
<td>50.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PRT002339</td>
<td>caspase-10</td>
<td>Cell</td>
<td>0.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PRT002349</td>
<td>procaspase-10</td>
<td>Cell</td>
<td>25.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>PRT002352</td>
<td>Fas:FasL</td>
<td>Cell</td>
<td>0.0</td>
<td></td>
<td>false</td>
<td></td>
<td></td>
</tr>
<tr>
<td>№</td>
<td>Title</td>
<td>Reactions</td>
<td>Kinetic law</td>
<td>Comment</td>
<td>Database references</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>---------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>k2_1/k2_2</td>
<td>Fas + FasL -&gt; Fas:FasL</td>
<td>$k_{2.1} [\text{Fas}] [\text{FasL}] - k_{2.2} [\text{Fas:FasL}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>k2_3</td>
<td>Fas:FasL -&gt; Fas:FasL Trimer</td>
<td>$k_{2.3} [\text{Fas:FasL}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>k2_4/k2_5</td>
<td>FADD + Fas:FasL Trimer -&gt; Fas:FasL Trimer:FADD</td>
<td>$k_{2.4} [\text{Fas:FasL Trimer}] [\text{FADD}] - k_{2.5} [\text{Fas:FasL Trimer:FADD}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>k2_15/k2_16</td>
<td>procaspase-8 +</td>
<td>$k_{2.15} [\text{procaspase-8}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fas:FasL Trimer:FADD:pro8:pro10 -&gt;</td>
<td></td>
<td>$[\text{Fas:FasL Trimer:FADD:pro8:pro10}] - k_{2.16} [\text{Fas:FasL Trimer:FADD:pro8:pro10}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>k2_17</td>
<td>Fas:FasL Trimer:FADD:(pro8)_2:pro10 -&gt; Fas:FasL Trimer:FADD:(p43/p41)_2:pro10</td>
<td>$k_{2.17}$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>k2_8/k2_9</td>
<td>FLIP long + Fas:FasL Trimer:FADD</td>
<td>$k_{2.8} [\text{FLIP long}] [\text{Fas:FasL Trimer:FADD}] - k_{2.9} [\text{Fas:FasL Trimer:FADD:FLIP}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>k2_12/k2_13</td>
<td>FLIP long + Fas:FasL Trimer:FADD:pro8:pro10 -&gt; Fas:FasL Trimer:FADD:FLIP</td>
<td>$k_{2.12} [\text{FLIP long}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>k2_14</td>
<td>Fas:FasL Trimer:FADD:FLIP:pro8:pro10 -&gt; Blocked p43/p41 + Blocked FASL:FAS Trimer:FADD</td>
<td>$k_{2.14}$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>k2_24</td>
<td>caspase-8 -&gt;</td>
<td>$k_{2.24} [\text{caspase-8}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>k2_19</td>
<td>Fas:FasL Trimer:FADD + FLIP short -&gt; Blocked FASL:FAS Trimer:FADD</td>
<td>$k_{2.19} [\text{FLIP short}] [\text{Fas:FasL Trimer:FADD}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>k2_20</td>
<td>Fas:FasL Trimer:FADD:pro8:pro10 + FLIP short -&gt; Blocked FASL:FAS Trimer:FADD</td>
<td>$k_{2.20} [\text{FLIP short}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>k2_6/k2_7</td>
<td>procaspase-8 + procaspase-10 + FADD:pro8:pro10 + FADD:FLIP:pro8:pro10</td>
<td>$k_{2.6} [\text{procaspase-8}] [\text{procaspase-10}]$</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rules

<table>
<thead>
<tr>
<th>№</th>
<th>Title</th>
<th>Formula</th>
<th>Comment</th>
</tr>
</thead>
</table>
| 1 | math-equation_0            | \[ \text{pro8\_total} = [\text{procaspase-8}] + [\text{Fas:FasL Trimer: FADD:pro8:pro10}] + \]
|   |                            | \[ [\text{Fas:FasL Trimer: FADD:FLIP:pro8:pro10}] + 2 \times [\text{Fas:FasL Trimer: FADD:(pro8\_2:pro10}] \]                           |         |
| 2 | math-equation_1            | \[ \text{pro10\_total} = [\text{procaspase-10}] + [\text{Fas:FasL Trimer: FADD:pro8:pro10}] + \]
|   |                            | \[ [\text{Fas:FasL Trimer: FADD:FLIP:pro8:pro10}] + [\text{Fas:FasL Trimer: FADD:(pro8\_2:pro10}] \]                               |         |
| 3 | math-equation_2            | \[ \text{p43p41\_total} = [\text{Blocked p43/p41}] + 2 \times [\text{Fas:FasL Trimer: FADD:(p43/p41\_2:p47/p43}] \]                |         |
| 4 | math-equation_3            | \[ \text{FLIPL\_total} = [\text{FLI long}] + [\text{Fas:FasL Trimer: FADD:FLIP}] + [\text{Fas:FasL Trimer: FADD:FLIP:pro8:pro10}] \] |         |

Differential equation system

1. \[ \frac{d\text{[Fas]}}{dt} = -k_{2\_1} \times [\text{Fas}] \times [\text{FasL}] + k_{2\_2} \times [\text{Fas:FasL}] \]
2. \[ \frac{d\text{[FasL]}}{dt} = -k_{2\_1} \times [\text{Fas}] \times [\text{FasL}] + k_{2\_2} \times [\text{Fas:FasL}] \]
3. \[ \frac{d\text{[caspase-8]}}{dt} = -k_{2\_24} \times [\text{caspase-8}] + k_{2\_18} \times [\text{Fas:FasL Trimer: FADD:(p43/p41\_2:p47/p43}] \]
4. \[ \frac{d\text{[procaspase-8]}}{dt} = -k_{2\_15} \times [\text{procaspase-8}] \times [\text{Fas:FasL Trimer: FADD:pro8:pro10}] + k_{2\_16} \times [\text{Fas:FasL Trimer: FADD:FLIP}] + k_{2\_7} \times [\text{procaspase-10}] + k_{2\_11} \times [\text{procaspase-10}] + k_{2\_6} \times [\text{procaspase-8}] \times [\text{Fas:FasL Trimer: FADD:FLIP}] + k_{2\_10} \times [\text{procaspase-8}] \times [\text{Fas:FasL Trimer: FADD:FLIP}] \]
5. \[ \frac{d\text{[FADD]}}{dt} = -k_{2\_4} \times [\text{Fas:FasL Trimer}] \times [\text{FADD}] + k_{2\_5} \times [\text{Fas:FasL Trimer: FADD}] \]
6. \[ \frac{d\text{[caspase-10]}}{dt} = k_{2\_18} \times [\text{Fas:FasL Trimer: FADD:(p43/p41\_2:p47/p43}] - k_{2\_23} \times [\text{caspase-10}] \]
7. \[ \frac{d\text{[procaspase-10]}}{dt} = -k_{2\_6} \times [\text{procaspase-8}] \times [\text{procaspase-10}] \times [\text{Fas:FasL Trimer: FADD}] + k_{2\_7} \times [\text{Fas:FasL Trimer: FADD:FLIP}] + k_{2\_11} \times [\text{procaspase-10}] + k_{2\_11} \times [\text{procaspase-10}] + k_{2\_6} \times [\text{procaspase-8}] \times [\text{Fas:FasL Trimer: FADD:FLIP}] + k_{2\_10} \times [\text{procaspase-8}] \times [\text{Fas:FasL Trimer: FADD:FLIP}] \]
8. \[ \frac{d\text{[Fas:FasL]}}{dt} = k_{2\_1} \times [\text{Fas}] \times [\text{FasL}] - k_{2\_2} \times [\text{Fas}] \times [\text{FasL}] - k_{2\_3} \times [\text{Fas}] \times [\text{FasL}] \]
Parameters fitting
Multi-experiment fitting
- Experiment types: time courses and steady states.
- Experimental data:
  - exact values of substance concentrations;
  - percentage values (relative to initial or completion values).

Distance functions:
mean, mean square and standard deviation weight methods.

Optimization methods:
- Adaptive simulated annealing;
- Cellular genetic algorithm;
- Evolution strategy with stochastic ranking (SRES);
- Particle swarm optimization;
- Deterministic global optimization method;
- Quadratic Hill-climbing.

ODE solvers:
- JVODE – ported to Java version of CVODE;
- Dormand-Prince (explicit Runge-Kutta scheme);
- Imex (implicit–explicit Runge–Kutta scheme);

Powerful user interface
Parameters optimization using Java Script
Parameters fitting – user interface

ID: Evolution strategy (SRES)

Description:

Stochastic ranking evolution strategy (SRES)\(^1\)

In the (\(\mu, \lambda\))-ES algorithm, the individual \(i\) is a set of real-vectors \((x_i, q_i)\), \(\forall \ i \in \{1, ..., \lambda\}\). The initial population is generated according to a uniform \(n\)-dimensional probability distribution over the search space \(S\). Let \(\delta x\) be an approximately measure of the expected distance to the global optimum; the initial setting for the "mean step sizes" should be...
CD95L module and results of fitting its dynamics to experimental data

Bentele M, 2004

Neumann L, 2010
<table>
<thead>
<tr>
<th>Method</th>
<th>BioUML (4 cores)</th>
<th>BioUML (1 core)</th>
<th>COPASI (1 core)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolutionary Programming</td>
<td>–</td>
<td>–</td>
<td>1 min 58,2sec</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 min 31,3 sec</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 min 16,6 sec</td>
</tr>
<tr>
<td>Particle swarm</td>
<td>7,1 sec</td>
<td>22,4 sec</td>
<td>1 min 32 sec</td>
</tr>
<tr>
<td></td>
<td>7,7 sec</td>
<td>15,3 sec</td>
<td>1 min 26,4 sec</td>
</tr>
<tr>
<td></td>
<td>6,9 sec</td>
<td>22,5 sec</td>
<td>1 min 07,1 sec</td>
</tr>
<tr>
<td>Stochastic Ranking Evolution Strategy</td>
<td>7,5 sec</td>
<td>23,4 sec</td>
<td>1 min 25,0 sec</td>
</tr>
<tr>
<td></td>
<td>7,47 sec</td>
<td>23,5 sec</td>
<td>1 min 5,6 sec</td>
</tr>
<tr>
<td></td>
<td>6,9 sec</td>
<td>22,2 sec</td>
<td>1 min 8,8 sec</td>
</tr>
<tr>
<td>Cellular genetic algorithm</td>
<td>7,7 sec</td>
<td>25,5 sec</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>7,5 sec</td>
<td>22,1 sec</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7,2 sec</td>
<td>20,8 sec</td>
<td></td>
</tr>
</tbody>
</table>
Standard supports:

SBML, SBGN, BioPAX, SED-ML, SBO, MIRIAM, CellML
SBML support

- import/export - level 1, 2, 3 (core)
- passed all tests from SBML test suite version 2.0.0 beta 1 (2010, April 2)
- extensions:
  - CellDesigner
  - SBGN-PD (own XML format)
- Biomodels
  - full text search
  - SBGN-PD (parse RDF annotation do determine specie types)
  - layout algorithms
  - on-line simulation
- Panther DB
  - full text search
  - reads CellDesigner extensions
  - SBGN-PD
CellDesigner support – comparison for 165 images from Panther DB and generated by BioUML
SBGN support

• diagram types: PD – beta; ER, AF – alpha
• defined as XML graphic notation for BioUML
• graphic notation can be edited using BioUML graphic notation editor
• special extension for SBML
• Reactome - SBGN-PD
  – full text search
  – SQL version is used
  – read diagram layout from Reactome
• BioPAX – SBGN-PD
  – level 2 (beta), level 3 (alpha)
  – auto layout
• TRANSPATH – SBGN-PD
  – full text search
  – auto-layout
BioPAX – import dialog
BioPAX example (BioCyc)
Glycolysis / Gluconeogenesis

Complete name: databases/KEGG
/Diagrams/map00010.xml
SED-ML support

- SED-ML import
  - only SBML models are supported now
  - stored in user’s project
- SED-ML changes presented as BioUML states
- SED-ML presented as workflow
- automated hierarchic layout
- on-line simulation
Simulation Experiment Description Markup Language (SED-ML) is an XML format for encoding simulation experiments, following the requirements defined by the MIBSE guidelines.

It allows the definition of the model(s) to be used, the experimental task(s) to be performed, and the result(s) to be produced. SED-ML currently only covers some of the elements for MIBSE compliant simulation descriptions. On the other hand, the XML file format can be extended by adding new elements to allow additional information to be shared with other tools. SEAMS is a SED-ML editor that provides a graphical interface to facilitate the creation of SED-ML files.
The Simulation Experiment Description Markup Language (SED-ML) is an XML-based format for encoding simulation experiments, following the requirements defined in the MAASE guidelines.

SED-ML allows the definition of the model(s) to be used, the experimental task(s) to be run, and the result(s) to be produced. SED-ML currently only covers some of the requirements for MAASE compliant simulation descriptions. On the other hand, the formalisation of SED-ML forces the information to be more precise than the bare MAASE requirements.
Systems biology:

reproducible

high-throughput data analyses

- analyses: algorithms, scripts, workflows
- integration with R/Bioconductor, Galaxy
- data: microarrays, NGS, ChIP-SEQ
- visualization: genome browser
Bbowtie
Alignment of short reads

Parameters:
- Input sequences
- Input sequences (fastq files)
- Species
- Number of mismatches
- Seed length
- Tole error
- No MAC round
- Max backtracks
- All alignments
- Mm alignments per read
- Max alignments per read
- All alignments
- Output track

Start page

Run
ID: Compute differentially expressed genes (Affymetrix probes)
Title: Compute differentially expressed genes (Affymetrix probes)
Size: 115
Complete name: analyses/Workflows/Detect differentially expressed genes/Compute differentially expressed genes (Affymetrix probes)

Description:
This workflow is designed to analyze datasets with more than two samples. For analysis of smaller number of samples please use the workflow: Compute differentially expressed genes using Hypergeometric test.
ID: Fold-Change calculation
Title: Fold-Change calculation
Size: 3
Attributes:
- `analysisName`: Fold-Change calculation
- `innerNodesPortFinder`: true
- `parameter.controlData`: ['data/Projects\test/Data/Psoriasis/Psoriasis_example2/Control normalized (MA55)/[Wall columns],0], ['GSM372316.CEL,COL',0], ['GSM372318.CEL,COL',0], ['GSM372316.CEL,COL',0]
Convert table identifiers using BioHub(s)

This analysis allows you to change the type of identifiers in the table and convert rows accordingly using chain of BioHubs. BioHub is an atomic converter capable to convert between two or more types (for example, convert "Genes: Ensembl" into "Proteins: Ensembl"). If direct conversion between two given types is impossible, this analysis will create the optimal chain of several BioHubs and use them subsequently.

Note that several non-trivial situations might occur during conversion:

- Single source ID matches to several target IDs. In this case source row will be copied several times, one copy per one target ID.
- Source ID doesn't match to any target ID. In this case source row will be copied zero times.
NGS

- интегрированные в BioUML методы (Bowtie, MACS, ChIPHorde, ChIPMunk, …)
- программы, интегрированные в Galaxy
- пакеты R
- аннотация найденных пиков (SNP, сайтов и т.п.)
- визуализация
- workflows
  - ChIP-SEQ
  - RNA-SEQ
- сборка и аннотация генома человека (в процессе)
- поддержка распараллеливания внешних программ как часть workflow
- база данных GTRD (на основе данных ChIP-SEQ)
- выделенные сервера
  - Amazon EC2 – по запросу
  - Biodatomatics – 64 ядра, 256 Гб памяти.
**ID:** Track to gene set

**Complete name:** analyses/Methods/BSA/Track to gene set

**Description:** Convert track to gene set (get all genes which lie near the sites from the track)
<table>
<thead>
<tr>
<th>siteID</th>
<th>Sequence (chromosome) name</th>
<th>type</th>
<th>from</th>
<th>to</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>unsure</td>
<td>4406941</td>
<td>4406954</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>unsure</td>
<td>4483877</td>
<td>4483876</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>unsure</td>
<td>4513053</td>
<td>4513076</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>unsure</td>
<td>4604251</td>
<td>4604259</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>unsure</td>
<td>4651054</td>
<td>4651065</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>unsure</td>
<td>4681953</td>
<td>4681998</td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>unsure</td>
<td>4775716</td>
<td>4775717</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>unsure</td>
<td>4792791</td>
<td>4792793</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>unsure</td>
<td>4800481</td>
<td>4800481</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>unsure</td>
<td>4843780</td>
<td>4843888</td>
<td>109</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>unsure</td>
<td>4847702</td>
<td>4847702</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>unsure</td>
<td>5060466</td>
<td>5060466</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>unsure</td>
<td>5073581</td>
<td>5073581</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>unsure</td>
<td>5222283</td>
<td>5222289</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>unsure</td>
<td>5776049</td>
<td>5776056</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>unsure</td>
<td>6005954</td>
<td>6005959</td>
<td>6</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>unsure</td>
<td>6107803</td>
<td>6107811</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>unsure</td>
<td>6204574</td>
<td>6204573</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>unsure</td>
<td>6349085</td>
<td>6349103</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>unsure</td>
<td>6438860</td>
<td>6438859</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>unsure</td>
<td>6448332</td>
<td>6448332</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>unsure</td>
<td>6861827</td>
<td>6861834</td>
<td>8</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>unsure</td>
<td>6921180</td>
<td>6921223</td>
<td>44</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>unsure</td>
<td>7037431</td>
<td>7037441</td>
<td>11</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>unsure</td>
<td>7079191</td>
<td>7079203</td>
<td>13</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>unsure</td>
<td>7114880</td>
<td>7114881</td>
<td>2</td>
</tr>
<tr>
<td>ID</td>
<td>Gene symbol</td>
<td>GSM288350: Structure</td>
<td>GSM288350: Schematic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000021470</td>
<td>0610007P08Rik</td>
<td>Intron x 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000042209</td>
<td>0610013F05Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000046727</td>
<td>061001012Rik</td>
<td>Intron x 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000025731</td>
<td>0610011F06Rik</td>
<td>5'; Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000027629</td>
<td>0610011L14Rik</td>
<td>Intron x 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000055312</td>
<td>0610012H03Rik</td>
<td>Intron x 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000058812</td>
<td>0610039K10Rik</td>
<td>5'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000085646</td>
<td>0610040B09Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000087341</td>
<td>0610040F04Rik</td>
<td>Intron x 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000060512</td>
<td>0610040J01Rik</td>
<td>Intron x 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000087361</td>
<td>0610043K17Rik</td>
<td>Intron x 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000087274</td>
<td>1010001N08Rik</td>
<td>5'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000071456</td>
<td>1110002L01Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000037822</td>
<td>1110003E01Rik</td>
<td>Exon; Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000022972</td>
<td>1110004E09Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000090266</td>
<td>1110005A03Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000027637</td>
<td>1110008F13Rik</td>
<td>Exon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000021023</td>
<td>1110008L16Rik</td>
<td>Intron x 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000052105</td>
<td>1110012J17Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000024737</td>
<td>1110014N23Rik</td>
<td>Exon; 3'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000042350</td>
<td>1110013G07Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000021482</td>
<td>1110013J18Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000022253</td>
<td>1110020G09Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000037455</td>
<td>1110021L09Rik</td>
<td>5'; Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000025169</td>
<td>1110031I12Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSMUSG00000043461</td>
<td>1110032A04Rik</td>
<td>Intron</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Фрагмент сценария для автоматической обработки ChIP-Seq данных
Фрагмент сценария для автоматической обработки ChIP-Seq данных
Galaxy – analyses methods

This tool uses Weblogo3 in Galaxy to generate a sequence logo. The input file must be a fasta file in your current history. It is recommended for (eg) viewing multiple sequence alignments output from the clustalw tool - set the output to fasta and feed it in to this tool.

A typical output looks like this

```
... image: /static/images/rgWebLogo3_test.jpg
....
```

**Note**

**Warning about input Fasta format files**
Galaxy - workflow
JavaScript host objects allows to merge R/Bioconductor and Java/BioUML worlds

R world

```
var r = R.rserve();
var pcr = data.get("data/Collaboration/test/Data/table1");
r,assignObject("pcr", pcr);
#R(rplot(r, R))
correlation <- cor(pcr, use='pairwise.complete.obs');
plot(hclust(as.dist(1-correlation), method='single'));
#end
```
```r
var r = R.rserve();
var pcr = data.get("data/Collaboration/test/Data/table1");
r.assignObject("pcr", pcr);
#R(rplot(r, R))
correlation <- cor(pcr, use='pairwise.complete.obs');
plot(hclust(as.dist(1-correlation), method='single'));
#end
```
/*

Чувствительность клеточных линий к лекарственным препаратам.

Для выбранного лекарственного препарата для каждой клеточной линии на графике показывается среднее значение IC50 и доверительный интервал 95%.
*/

var r = R.rserve();
var drugs = data.get("data/cancer\ therapy/data/drugs/moscow.drugs.txt");
var ci95 = data.get("data/cancer\ therapy/data/drugs/moscow.drugs.ci95.txt");
r,assignObject("drugs", drugs);
r,assignObject("ci95", ci95);

#R(rplot(r, R))
drug <- 'Platina'
x <- drugs[,drug]
ci95.from <- ci95[!is.na(x), paste(drug,'from',sep='_.')]
ci95.to <- ci95[!is.na(x), paste(drug,'to', sep='_.')]
cells <- rownames(drugs)[!is.na(x)]
x <- x[!is.na(x)]
plot(0, 0, xlab='Cell line', ylab='IC50, 95% CI', main=drug, ylim=c(min(ci9)
axis(1, 1:length(cells), cells)
arrows(1:length(x), ci95.from, 1:length(x), ci95.to, angle=90, code=3, leng
#end

opening rplot...
*/

Чувствительность клеточных линий к лекарственным препаратам.

Для выбранного лекарственного препарата для каждой клеточной линии на графике показывается среднее значение IC50 и доверительный интервал 95%.

```r
var r = R.rserve();
var drugs = data.get("data/cancer therapy/data/drugs/moscow.drugs.txt");
var ci95 = data.get("data/cancer therapy/data/drugs/moscow.drugs.ci95.txt");
r.assignObject("drugs", drugs);
```

**Platina**

IC50, 95% CI

<table>
<thead>
<tr>
<th>Cell line</th>
<th>IC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>A549</td>
<td>2</td>
</tr>
<tr>
<td>h1299</td>
<td>4</td>
</tr>
<tr>
<td>h358</td>
<td>6</td>
</tr>
<tr>
<td>h23</td>
<td>8</td>
</tr>
<tr>
<td>h460</td>
<td>10</td>
</tr>
<tr>
<td>h322</td>
<td>2</td>
</tr>
<tr>
<td>h292</td>
<td>4</td>
</tr>
<tr>
<td>calu1</td>
<td>6</td>
</tr>
</tbody>
</table>
void boxPlot(String[] columns, Double[] values1, Double[] values2...)

void boxPlot(TableDataCollection table, String columnName1, String columnName2...)

Create and show box and whisker chart

Parameters(1):

- columns - column names
- values1 - values for column
- values2... - values for column

Parameters(2):

ID: boxPlot
Description: Create and show box and whisker chart
**ID:** boxPlot  
**Description:** Create and show box and whisker chart

```javascript
boxPlot(["c1","c2","c3"],[1,2,3],[1,1.5,3],[5,2,3,2]);
```
JavaScript function

ru.biosoft.plugins.jri.RObject rserve()

returns RObject for RServe type of R connection

Parameters:

Returns:

RServe R object
JavaScript host object

ScriptableObject R

Facade for R usage

Functions:

- `ru.biosoft.plugins.jri.RObject connect(String host, String port)`
- `ru.biosoft.plugins.jri.RObject local()`
- `ru.biosoft.plugins.jri.RObject rserve()`

```javascript
js> r=R.rserve();
Rserve is running.
```

R Serve R object
R> hist(rnorm(50))

$breaks
[1] -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
[9] 2.0

$counts
[1] 2 5 5 9 13 9 5 2

$intensities
[1] 0.08 0.2 0.2 0.36 0.52 0.36 0.2 0.08

$density
[1] 0.08 0.2 0.2 0.36 0.52 0.36 0.2 0.08

$mids
[1] -1.75 -1.25 -0.75 -0.25 0.25 0.75 1.25 1.75

$xname
rnorm(50)

$equidist
org.rosuda.REngine.REXPLogical@1b5b4729

opening rplot...
Opening rplot...

R> hist(rnorm(50))

$breaks
[1] -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
[9] 2.0

$count
[1] 2 5 5 9 13 9 5 2

$intensities
[1] 0.08 0.2 0.2 0.36 0.52 0.36 0.2 0.08

$density
[1] 0.08 0.2 0.2 0.36 0.52 0.36 0.2 0.08

$mids
[1] -1.75 -1.25 -0.75 -0.25 0.25 0.75 1.25 1.75

$xname
rnorm(50)

$equidist
org.rosuda.REngine.REXPLogical@1b5b4729[1]

Opening rplot...
R> hist(rnorm(50))

$breaks
 [1] -2.0 -1.5 -1.0 -0.5  0.0  0.5  1.0  1.5  2.0

$counts
 [1] 2 5 5 9 13 9 5

$intensities
 [1] 0.08 0.2 0.2 0

$density
 [1] 0.08 0.2 0.2 0

$mids
 [1] -1.75 -1.25 -0.75 -0.25  0.25  0.75  1.25  1.75

$xname
 rnorm(50)

hist(rnorm(50))
Computational modeling

Experimental data

Experimental hypothesis validation

Data repository for collaborative work

Data Analysis

COMBINE Systems Biology
Systems Biology

- Computational modeling
- Experimental data
- Experimental hypothesis validation
- Data repository for collaborative work

Data Analysis

LIMS
BioUML – LIMS
(Laboratory Information Management System)

– is being developed by Data Integrated Solutions Inc. (Pittsburg, USA)
  start-up organized by Maxim Mikheev and Fedor Kolpakov
– experiment is represented as workflow
– automated management by laboratory equipment
– automated generation of tasks for laboratory staff
– automated upload of experimental data into repository
– automated row data analysis
Genome browser
Genome browser: main features

- uses AJAX and HTML5 `<canvas>` technologies
- interactive - dragging, semantic zoom
- tracks support
  - Ensembl
  - DAS-servers
  - user-loaded BED/GFF/Wiggle files
site

siteID: 463
Sequence name: 21
type: q22.12
from: 34700001
to: 36700000
length: 2000000
The **Distributed Annotation System** (DAS) defines a communication protocol used to exchange annotations on genomic or protein sequences.

It is motivated by the idea that such annotations should not be provided by single centralized databases, but should instead be spread over multiple sites. Data distribution, performed by DAS servers, is separated from visualization, which is done by DAS clients.

DAS is a client-server system in which a single client integrates information from multiple servers. It allows a single machine to gather up sequence annotation information from multiple distant web sites, collate the information, and display it to the user in a single view.

DAS is heavily used in the genome bioinformatics community. Over the last years we have also seen growing acceptance in the protein sequence and structure communities.
Main commands for DAS servers

**Meta information**
- sources - list of data sources that are available from this server
- entry points - list of reference objects known by a data source
- types* - types of annotation available for a data source
- stylesheet - to retrieve the server's recommendations on formatting annotations retrieved from it.

**Protein (DNA) data**
- sequence
- feature* - the annotations available for a reference segment
  - positional
  - non positional
- structure - a protein 3D structure, including metadata and coordinates

**Extensions**
- interactions
- alignment
- volpmap

* используются онтологии (см. далее)
Often, different DAS annotation sources provide similar features. It is useful for client software to have a way to formally categorise these annotations so that they can be presented to the user in an intelligent manner. This also allows a user to more easily access the types of data they are interested in, which is especially important given the now large number of DAS servers. To facilitate this, the protocol now incorporates formal ontologies.

- **Sequence Types and Features** (namespace: SO) contains terms for both genomic and protein sequence annotations.

- **Protein Modifications** (namespace: MOD) contains terms for post-translational modifications.

- **BioSapiens Annotations** (namespace: BS) originally developed for the BioSapiens consortium, this ontology contains protein-focussed terms for nonpositional annotations (e.g. publications).
### Features

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Description</th>
<th>Expression</th>
<th>Visible</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS-00138</td>
<td>DescribedString</td>
<td>BS-00138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS-01019</td>
<td>DescribedString</td>
<td>BS-01019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS-01040</td>
<td>DescribedString</td>
<td>BS-01040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO:0000419</td>
<td>GenomeBrowser</td>
<td>SO:0000419</td>
<td>bsa.createProject/data</td>
<td></td>
</tr>
<tr>
<td>SO:0000039</td>
<td>GenomeBrowser</td>
<td>SO:0000039</td>
<td>bsa.createProject/data</td>
<td></td>
</tr>
<tr>
<td>id</td>
<td>Text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sequences_1</td>
<td>DescribedString</td>
<td>sequences</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Showing 1 to 7 of 7 entries
Protein/gene set annotation (non positional features)
Protein/gene set annotation (non positional features)

A6NKL6
BS:00138 (Nusbaum C. (2005), Pickard B.S. (2005.))

View in PubMed
Sequence view

ID: uniprot-hs-chr18
Size: 264
Complete name: data/Collaboration/test/Data/Proteome/uniprot-hs-chr18

sequences_1 (Length: 621)

1. MIATGLLKL1SARKODPLRPSCFIPKKERKAKKRRDNDVVVV0RKLKLC515GLTIALG1 60
   LVLLCVIAMAUVGWYFKATCNR3EGQKLPFASPSSHVPPTANN33SESGHRSHBFRAP 120
   LGVSSSAGARPSTPFAAASPSSTTSVF3IFIGLSYKLKFGELINGILG1FLF 180
   101. ICANVINGHNRDARKT1NINRLDLY3TVDVHHRLAKCDDAAA5AAAASSSSSAAAAPA 240
   241. PPGAPIFLNGFLVQDS6GLFEGGCDGAAMALSKSVPSHAPSSGSGRPGAA 300
   301. SFFDLASAFSCPFNEPFPSITAIYVSYRERSGYASRRAAATAAAAASSSCSCPAMEPPE 360
   351. SWQRCTQATSSVFESLSAPALLFLQGRDCGDAGADACSSWQFQERGQIFQROI 420
   421. SMHQLGAEAGGSGARHGPRKEPFAQGAAAREKRGQGGLRPFTGRYAILRER3TSSGLPDYRA 480
   481. PPSEPFFPF0SADPS3PLAASPSFPLKEGIPTFTRSDS250DFDS2MNKTRPFL 540
   541. REIEQIESVQDVAQTFQRIKVR1APFUGAEGSSPSGRGQEPFTATEPQVQRFQFNT 600
   601. LMT13SKMAHVEEELSTGI 621
<table>
<thead>
<tr>
<th>ID</th>
<th>Gene symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSMUSG00000017470</td>
<td>0610007P08Rik</td>
</tr>
<tr>
<td>ENSMUSG00000022009</td>
<td>0610013F05Rik</td>
</tr>
<tr>
<td>ENSMUSG00000017470</td>
<td>0610010101Rik</td>
</tr>
<tr>
<td>ENSMUSG00000025731</td>
<td>0610011F06Rik</td>
</tr>
<tr>
<td>ENSMUSG00000027629</td>
<td>0610011L04Rik</td>
</tr>
<tr>
<td>ENSMUSG00000055312</td>
<td>0610012H03Rik</td>
</tr>
<tr>
<td>ENSMUSG00000058312</td>
<td>0610039K10Rik</td>
</tr>
<tr>
<td>ENSMUSG00000085646</td>
<td>0610040B09Rik</td>
</tr>
<tr>
<td>ENSMUSG00000087341</td>
<td>0610040F04Rik</td>
</tr>
<tr>
<td>ENSMUSG00000060512</td>
<td>0610040J01Rik</td>
</tr>
<tr>
<td>ENSMUSG00000087361</td>
<td>0610043K17Rik</td>
</tr>
<tr>
<td>ENSMUSG00000087274</td>
<td>1010001N08Rik</td>
</tr>
<tr>
<td>ENSMUSG000000711456</td>
<td>1110002L01Rik</td>
</tr>
<tr>
<td>ENSMUSG00000037822</td>
<td>1110003E01Rik</td>
</tr>
<tr>
<td>ENSMUSG00000022972</td>
<td>1110004E09Rik</td>
</tr>
<tr>
<td>ENSMUSG00000090266</td>
<td>1110005A03Rik</td>
</tr>
<tr>
<td>ENSMUSG00000072637</td>
<td>1110008F13Rik</td>
</tr>
<tr>
<td>ENSMUSG00000021023</td>
<td>1110008L16Rik</td>
</tr>
<tr>
<td>ENSMUSG00000052105</td>
<td>1110012J17Rik</td>
</tr>
<tr>
<td>ENSMUSG00000024757</td>
<td>1110014N23Rik</td>
</tr>
<tr>
<td>ENSMUSG00000042950</td>
<td>1110015G07Rik</td>
</tr>
<tr>
<td>ENSMUSG00000021482</td>
<td>1110018L18Rik</td>
</tr>
<tr>
<td>ENSMUSG00000022253</td>
<td>1110020L09Rik</td>
</tr>
<tr>
<td>ENSMUSG00000032545</td>
<td>1110021L09Rik</td>
</tr>
<tr>
<td>ENSMUSG00000025169</td>
<td>111003102Rik</td>
</tr>
<tr>
<td>ENSMUSG00000043461</td>
<td>1110032A04Rik</td>
</tr>
</tbody>
</table>
3D view for protein structures from PDB
Chemoinformatics
CDK – Chemical development kit
PASS: Computerized Prediction of Biological Activity Spectra for Chemical Substances developed by group of Prof. Vladimir Poroikov, IBMC, Moscow
<table>
<thead>
<tr>
<th>ID</th>
<th>No</th>
<th>Group</th>
<th>Number</th>
<th>IEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N-acetylneuraminy)-galactosylglucosylceramide</td>
<td>1</td>
<td>M</td>
<td>6</td>
<td>5.079</td>
</tr>
<tr>
<td>N-acetylgalactosaminyltransferase inhibitor</td>
<td>2</td>
<td>M</td>
<td>97</td>
<td>0.028</td>
</tr>
<tr>
<td>3 Beta-hydroxy-delta 5-steroid dehydrogenase inhibitor</td>
<td>3</td>
<td>M</td>
<td>1339</td>
<td>1.719</td>
</tr>
<tr>
<td>5 Lipoxygenase inhibitor</td>
<td>4</td>
<td>M</td>
<td>1324</td>
<td>0.622</td>
</tr>
<tr>
<td>5-Alpha-reductase inhibitor</td>
<td>5</td>
<td>G</td>
<td>200</td>
<td>0.714</td>
</tr>
<tr>
<td>ABCA1 expression enhancer</td>
<td>6</td>
<td>M</td>
<td>24</td>
<td>1.565</td>
</tr>
<tr>
<td>ADAM10 endopeptidase inhibitor</td>
<td>7</td>
<td>M</td>
<td>441</td>
<td>0.693</td>
</tr>
<tr>
<td>ADP ribose polymerase 1 inhibitor</td>
<td>8</td>
<td>M</td>
<td>757</td>
<td>1.157</td>
</tr>
<tr>
<td>ADP ribose polymerase inhibitor</td>
<td>9</td>
<td>M</td>
<td>30</td>
<td>4.906</td>
</tr>
<tr>
<td>AICAR transformylase inhibitor</td>
<td>10</td>
<td>M</td>
<td>960</td>
<td>1.091</td>
</tr>
<tr>
<td>AMPA receptor antagonist</td>
<td>11</td>
<td>M</td>
<td>25</td>
<td>6.151</td>
</tr>
<tr>
<td>ATM kinase inhibitor</td>
<td>12</td>
<td>M</td>
<td>78</td>
<td>1.050</td>
</tr>
<tr>
<td>ATPase (Vacuolar H+) inhibitor</td>
<td>13</td>
<td>M</td>
<td>216</td>
<td>3.661</td>
</tr>
<tr>
<td>Acetylcholine nicotinic antagonist</td>
<td>14</td>
<td>M</td>
<td>839</td>
<td>2.446</td>
</tr>
<tr>
<td>Adenosine A3 receptor agonant</td>
<td>15</td>
<td>M</td>
<td>90</td>
<td>0.197</td>
</tr>
<tr>
<td>Adenylyl cyclase inhibitor</td>
<td>16</td>
<td>M</td>
<td>132</td>
<td>5.756</td>
</tr>
<tr>
<td>Aggrecanase inhibitor</td>
<td>17</td>
<td>M</td>
<td>236</td>
<td>0.538</td>
</tr>
<tr>
<td>Alkylator</td>
<td>18</td>
<td>E</td>
<td>554</td>
<td>5.976</td>
</tr>
<tr>
<td>Alkylphospholipid</td>
<td>19</td>
<td>M</td>
<td>8</td>
<td>0.018</td>
</tr>
<tr>
<td>Aminopeptidase N inhibitor</td>
<td>20</td>
<td>M</td>
<td>131</td>
<td>2.064</td>
</tr>
</tbody>
</table>
**PASS analysis**

**ANTIVIRAL**

<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>ACTIVITY</th>
<th>PASS_ACTIVITY_SPECTRUM</th>
<th>PASS_MRNA_CUTOFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR1</td>
<td></td>
<td>Antineoplastic</td>
<td>0.142 0.063 Antimitotic, 0.142</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.125 Interleukin, 0.149</td>
<td>0.038 DNA damaging, 0.153</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.124 Interleukin 6 antagonist, 0.161</td>
<td>0.134 Lipocortin synthesis antagonist, ...</td>
</tr>
<tr>
<td>STR10</td>
<td></td>
<td></td>
<td>0.041 0.034 Protein kinase C zeta inhibitor, 0.056 0.054</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.066 0.050 Histamine agonist, 0.100 0.024 P-glycoprotein 1 inhibitor, 0.121 0.037 Tubulin antagonist, ...</td>
<td>34</td>
</tr>
<tr>
<td>STR100</td>
<td></td>
<td></td>
<td>0.031 0.017 CC chemokine 1 receptor antagonist, 0.063</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.035 Dopamine D2 agonist, 0.073 0.056 Mannosidase inhibitor, 0.084 0.032 5 Lipoxigenase inhibitor, 0.150</td>
<td>40</td>
</tr>
</tbody>
</table>

**CONVULSANT**

<table>
<thead>
<tr>
<th>Structure</th>
<th>ACTIVITY</th>
<th>PASS_ACTIVITY_SPECTRUM</th>
<th>PASS_MRNA_CUTOFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR10</td>
<td></td>
<td>0.041 0.034 Protein kinase C zeta inhibitor, 0.056 0.054</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.066 0.050 Histamine agonist, 0.100 0.024 P-glycoprotein 1 inhibitor, 0.121 0.037 Tubulin antagonist, ...</td>
<td>34</td>
</tr>
<tr>
<td>STR100</td>
<td></td>
<td>0.031 0.017 CC chemokine 1 receptor antagonist, 0.063</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.035 Dopamine D2 agonist, 0.073 0.056 Mannosidase inhibitor, 0.084 0.032 5 Lipoxigenase inhibitor, 0.150</td>
<td>40</td>
</tr>
</tbody>
</table>

**Columns (double-click to select):**

- Identifier
- Type
- Date
- Title
- Description

**Example:** Score > 0.5 & Group_number = 1
Acknowledgements

Part of this work was partially supported by the grant:
European Committee grant №037590 “Net2Drug”
European Committee grant №202272 “LipidomicNet”
Integration and interdisciplinary grants №16, 91 of SB RAS.

BioUML team

Software developers

Nikita Tolstyh
Tagir Valeev
Anna Ryabova

Ilya Kiselev
Elena Kutumova
Alexey Shadrin

Biologists

Ruslan Sharipov
Ivan Yevshin